

Coordinate Descent

Nipun Batra

IIT Gandhinagar

July 29, 2025

where:

$$\hat{y}_i^{(-j)} = \theta_0 x_i^0 + \dots + \theta_d x_i^d$$

is \hat{y}_i without θ_j

Coordinate Descent for Unregularized regression

$$\text{Set } \frac{\partial \text{RSS}(\theta_j)}{\partial \theta_j} = 0$$

$$\theta_j = \sum_{i=1}^N \frac{(y_i - (\theta_0 x_i^0 + \dots + \dots + \theta_d x_i^d)) (x_i^j)}{(x_i^j)^2} = \frac{\rho_j}{z_j}$$

$$\rho_j = \sum_{i=1}^N x_i^j (y_i - \hat{y}_i^{(-j)})$$

$$z_j = \sum_{i=1}^N (x_i^j)^2$$

z_j is the squared of ℓ_2 norm of the j^{th} feature

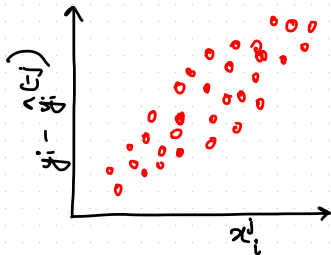
UNDERSTANDING J_j IN COORDINATE DESCENT

$$J_j = \sum_{i=1}^N x_i^j (y_i - \hat{y}_i^{(j)})$$

UNDERSTANDING p_j IN COORDINATE DESCENT

$$p_j = \sum_{i=1}^N x_i^j (y_i - \hat{y}_i^{(j)})$$

CASE 1

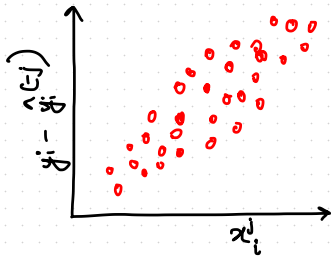


x_i^j STRONG +ve CORR.
WITH $y_i - \hat{y}_i^{(j)}$

UNDERSTANDING β_j IN COORDINATE DESCENT

$$\beta_j = \sum_{i=1}^N x_i^j (y_i - \hat{y}_i^{(-j)})$$

CASE 1



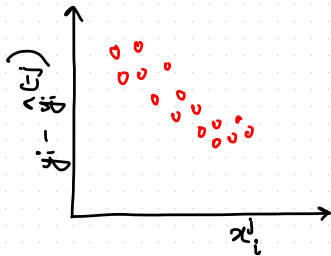
x_i^j STRONG +ve CORR.
WITH $y_i - \hat{y}_i^{(-j)}$

⇓
 j^{th} FEATURE IS IMPT.
AND ITS COEFFICIENT
+ve

UNDERSTANDING β_j IN COORDINATE DESCENT

$$\beta_j = \sum_{i=1}^N x_i^j (y_i - \hat{y}_i^{(-j)})$$

CASE 11



x_i^j STRONG -VE CORR.
WITH $y_i - \hat{y}_i^{(-j)}$

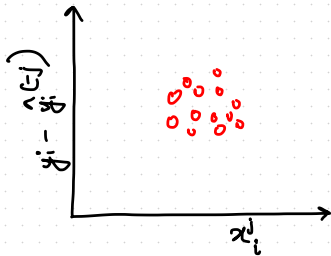


j^{th} FEATURE IS IMPT.
AND ITS COEFF. -VE

UNDERSTANDING P_j IN COORDINATE DESCENT

$$P_j = \sum_{i=1}^N x_i^j (y_i - \hat{y}_i^{(j)})$$

CASE II



x_i^j WEAK WITH $y_i - \hat{y}_i^{(j)}$ CORR.
↓
jth FEATURE IS **NOT** IMPT.
AND ITS COEF. $\rightarrow 0$

Coordinate Descent for Lasso Regression

$$\text{Minimize } \underbrace{\sum_{i=1}^N \epsilon^2 + \delta^2 \{|\theta_0| + |\theta_1| + \dots |\theta_j| + \dots |\theta_d|\}}_{\text{LASSO OBJECTIVE}}$$

$$\frac{\partial}{\partial \theta_j} (\text{LASSO OBJECTIVE}) = -2\rho_j + 2\theta_j z_j + \delta^2 \frac{\partial}{\partial \theta_j} |\theta_j|$$

$$\frac{\partial}{\partial \theta_j} |\theta_j| = \begin{cases} 1 & \theta_j > 0 \\ [-1, 1] & \theta_j = 0 \\ -1 & \theta_j < 0 \end{cases}$$

Coordinate Descent for Lasso Regression

► **Case 1:** $\theta_j > 0$

$$2\rho_j + 2\theta_j z_j + \delta^2 = 0$$

$$\theta_j = \frac{\rho_j - \frac{\delta^2}{2}}{z_j}$$

$$\rho_j > \frac{\delta^2}{2} \Rightarrow \theta_j = \frac{\rho_j - \frac{\delta^2}{2}}{z_j}$$

Coordinate Descent for Lasso Regression

► **Case 1:** $\theta_j > 0$

$$2\rho_j + 2\theta_j z_j + \delta^2 = 0$$

$$\theta_j = \frac{\rho_j - \frac{\delta^2}{2}}{z_j}$$

$$\rho_j > \frac{\delta^2}{2} \Rightarrow \theta_j = \frac{\rho_j - \frac{\delta^2}{2}}{z_j}$$

Coordinate Descent for Lasso Regression

► **Case 1:** $\theta_j > 0$

$$2\rho_j + 2\theta_j z_j + \delta^2 = 0$$

$$\theta_j = \frac{\rho_j - \frac{\delta^2}{2}}{z_j}$$

$$\rho_j > \frac{\delta^2}{2} \Rightarrow \theta_j = \frac{\rho_j - \frac{\delta^2}{2}}{z_j}$$

► **Case 2:** $\theta_j < 0$

$$\rho_j < \frac{\delta^2}{2} \Rightarrow \theta_j = \frac{\rho_j + \delta^2/2}{z_j} \tag{1}$$

Coordinate Descent for Lasso Regression

► **Case 3:** $\theta_j = 0$

$$\frac{\partial}{\partial \theta_j}(\text{LASSO OBJECTIVE}) = -2\rho_j + 2\theta_j z_j + \underbrace{\delta^2 \frac{\partial}{\partial \theta_j} |\theta_j|}_{[-1,1]}$$
$$\in \underbrace{[-2\rho_j - \delta^2, -2\rho_j + \delta^2]}_{\{0\} \text{ lies in this range}}$$

$$-2\rho_j - \delta^2 \leq 0 \text{ and } -2\rho_j + \delta^2 \leq 0$$
$$-\frac{\delta^2}{2} \leq \rho_j \leq \frac{\delta^2}{2} \Rightarrow \theta_j = 0$$

Summary of Lasso Regression

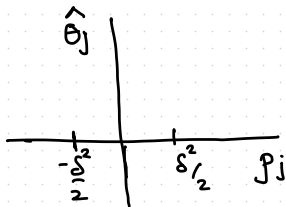
$$\theta_j = \begin{bmatrix} \frac{\rho_j + \frac{\delta^2}{2}}{z_j} & \text{if } \rho_j < -\frac{\delta^2}{2} \\ 0 & \text{if } -\frac{\delta^2}{2} \leq \rho_j \leq \frac{\delta^2}{2} \\ \frac{\rho_j - \frac{\delta^2}{2}}{z_j} & \text{if } \rho_j > \frac{\delta^2}{2} \end{bmatrix} \quad (2)$$

LASSO (SOFT) THRESHOLDING

$$\theta_j = \begin{cases} \frac{\beta_j + \delta^2/2}{z_j} & \text{if } \beta_j < -\delta^2/2 \\ 0 & \text{if } -\frac{\delta^2}{2} \leq \beta_j \leq \frac{\delta^2}{2} \\ \frac{\beta_j - \delta^2/2}{z_j} & \text{if } \beta_j > \delta^2/2 \end{cases}$$

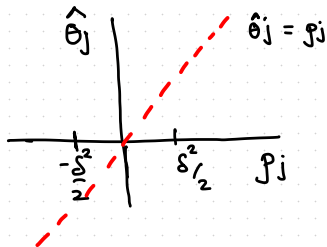
LASSO (SOFT) THRESHOLDING

$$\hat{\theta}_j = \begin{cases} \frac{p_j + \delta^2/2}{z_j} & \text{if } p_j < -\delta^2/2 \\ 0 & \text{if } -\frac{\delta^2}{2} \leq p_j \leq \frac{\delta^2}{2} \\ \frac{p_j - \delta^2/2}{z_j} & \text{if } p_j > \delta^2/2 \end{cases}$$



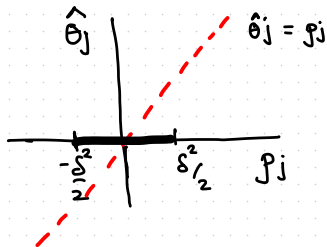
LASSO (SOFT) THRESHOLDING

$$\hat{\theta}_j = \begin{cases} \frac{p_j + \delta^2/2}{z_j} & \text{if } p_j < -\delta^2/2 \\ 0 & \text{if } -\frac{\delta^2}{2} \leq p_j \leq \frac{\delta^2}{2} \\ \frac{p_j - \delta^2/2}{z_j} & \text{if } p_j > \delta^2/2 \end{cases}$$



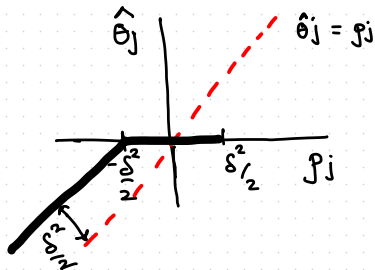
LASSO (SOFT) THRESHOLDING

$$\hat{\theta}_j = \begin{cases} \frac{p_j + \delta^2/2}{z_j} & \text{if } p_j < -\delta^2/2 \\ 0 & \text{if } -\delta^2/2 \leq p_j \leq \delta^2/2 \\ \frac{p_j - \delta^2/2}{z_j} & \text{if } p_j > \delta^2/2 \end{cases}$$



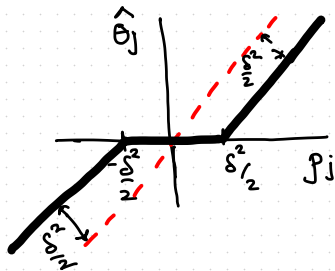
LASSO (SOFT) THRESHOLDING

$$\hat{\theta}_j = \begin{cases} \frac{p_j + \delta^2/2}{z_j} & \text{if } p_j < -\delta^2/2 \\ 0 & \text{if } -\frac{\delta^2}{2} \leq p_j \leq \frac{\delta^2}{2} \\ \frac{p_j - \delta^2/2}{z_j} & \text{if } p_j > \delta^2/2 \end{cases}$$



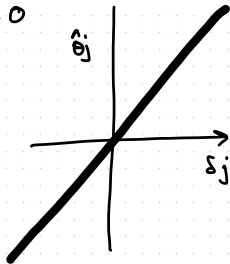
LASSO (SOFT) THRESHOLDING

$$\hat{\theta}_j = \begin{cases} \frac{p_j + \delta^2/2}{z_j} & \text{if } p_j < -\delta^2/2 \\ 0 & \text{if } -\frac{\delta^2}{2} \leq p_j \leq \frac{\delta^2}{2} \\ \frac{p_j - \delta^2/2}{z_j} & \text{if } p_j > \delta^2/2 \end{cases}$$



LASSO (SOFT) THRESHOLDING

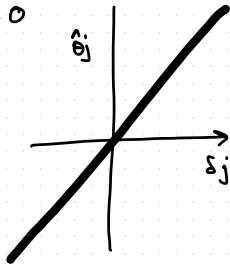
$$s^2 = 0$$



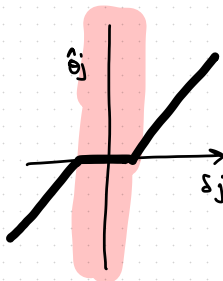
NO REGULARISATION

LASSO (SOFT) THRESHOLDING

$$s^2 = 0$$



NO REGULARISATION



REGULARISATION
↓
SPARSITY